Proceedings of the 2009 Winter Simulation Conference
M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin and R. G. Ingalls, eds.

REAL WORLD DATA COLLECTION FOR MODEL VALIDATION

Alan Cowdale

Thomson Building
RAF Waddington
Operational Analysis Element
Lincoln, LN5 9WA, UK.

ABSTRACT

The Operational Analysis Element of the Air Warfare Centre provides the scientific support to front line units of the Royal Air Force, and consequently is required to develop, maintain and use simulation models and tools for addressing a range of military problems. On occasions there will be opportunities to undertake controlled trials to collect data for model validation. While potentially immensely valuable for the validation of the simulation, trials can present their own challenges and analysis issues.

1 INTRODUCTION

The potential for Modeling and Simulation (M&S) to predict the performance of a system is increasingly being used to assist in making acquisition decisions, which in the Defense domain is called *simulation-based acquisition.*

M&S can be utilized throughout the acquisition process, initially in supporting the requirements definition process and continuing through to operational testing of the system and in-service evaluation. In the United Kingdom (UK), the Ministry of Defence (MOD) has developed an Integrated Test, Evaluation and Acceptance (ITEA) process for ensuring that a supplied solution meets the user's needs (MOD 2008) throughout the in-service life of the system. The ITEA process is being developed to encompass M&S and synthetic environments as well as the more traditional facilities and range capabilities associated with physical hardware testing. While the UK MOD ITEA approach is not a legal requirement (and hence is different to the Operational Test and Evaluation undertaken by the United States Department of Defense under Title 10 US Legal Code 2399) the process is becoming mandatory for all major UK Defence procurement projects.

Before M&S tools are used to support any key procurement or organizational decision making, it is essential however that the tools are accompanied by relevant verification, validation and, where appropriate, accreditation activities.

Planning, managing and conducting the verification and validation process requires a life-cycle approach that covers both the initial delivery and the subsequent development of the M&S tool.

The more successful the M&S tool becomes, the greater the importance of the verification and validation process. However, as the tool becomes more successful, the scope of the tool's usage often increases and with it the difficulties of ensuring adequate verification and validation.

The Operational Analysis Element (OAE) of the Air Warfare Centre (AWC) provides the scientific support to front line units of the Royal Air Force, and consequently is required to develop, maintain and use simulation models and tools for addressing a range of military problems. On occasions there will be opportunities to undertake controlled trials to obtain data that can be used to aid model validation. While immensely valuable for the validation of the simulation model, trials of this nature can however present their own challenges and analysis issues.

2 THROUGH-LIFE VALIDATION

Validation, in a general sense, is the process of assessing analytical capability against standards defining a fully credible system. For a simulation modeller, validation is establishing that the tools, data and models available for a study in a particular domain are capable of providing a credible simulation of the respective systems. This definition acknowledges that a set of validation criteria for the model alone is of limited value since there is almost always a clear linkage between the model and its data, with the latter very often needing to be tailored to reflect its intended use.

There are many books and papers dealing with methodologies for simulation validation (for example Law and Kelton 2000). The majority of the methodologies will invariably involve a combination of;

1. Conversations with subject matter experts.
2. Comparison with existing theory.
3. Comparison with observations of the current system (Lada et al. 2005).
4. Comparison with similar models and studies (Cowdale 2005).
5. Intuition.

The simulation model may well be developed in parallel with the acquisition project. Consequently the model will need to be continually reviewed to ensure that the validation has kept pace with the model development and the associated model usage.

Within the AWC, simulation models are generally used in one of two modes;

1. To support Developmental Testing – to ensure that the system under test meets the design (or contractual) specification.
2. To support Operational Testing – to ensure that the system under test meets the operational requirements, and that changes to doctrine and tactical procedures can be evaluated to improve system effectiveness.

Ideally the same model can be used in both modes (therefore reducing development costs and maintenance overheads), although the requirements for the different types of Testing may require different ranges of input variables and output products.

3 DATA COLLECTION FOR MODEL VALIDATION

The initial phase of the simulation validation process, for both Developmental and Operational Testing, will invariably involve running the simulation and analyzing the results (Figure 1). The results of the analysis are then used to either support validation of the model, or to identify areas of weakness within the simulation model or within the selected input data.

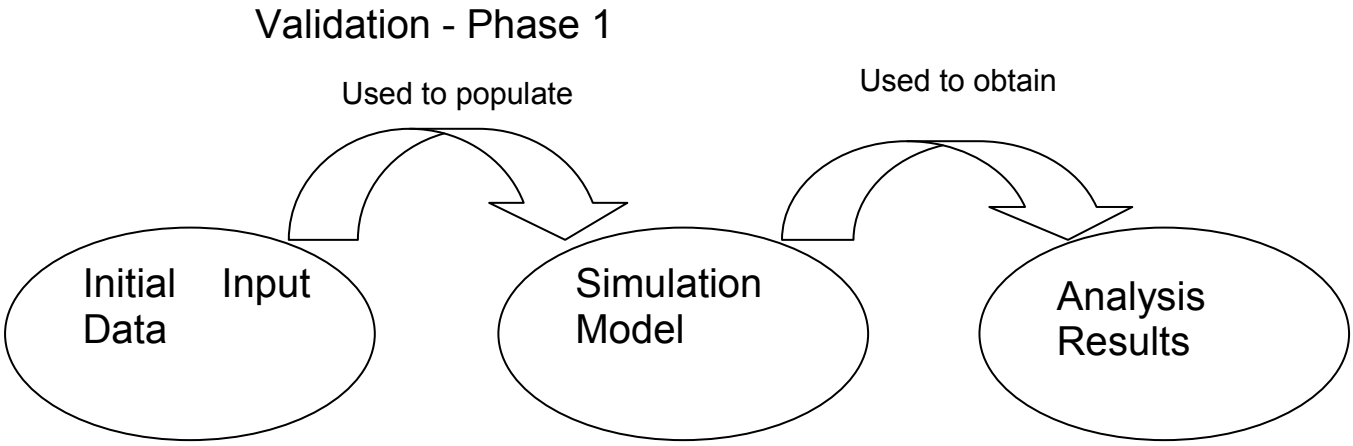

Figure 1: Initial Validation Process

Data validation is often not considered to be part of the model validation process because it is usually difficult, time consuming, and costly to obtain sufficient, accurate, and appropriate data (Sargent 2003). As a result, the lack of data validation is often the reason that attempts to validate a model as a whole fail. Even if high quality data is available, the complexity of the system being modeled may often make the data difficult to use in the model validation process (Cowdale 2006).

The simulation model will however only ever be a representation of the full system, and as such the key decision for the modeller and the customer is invariably 'is the model valid enough for the simulation study being undertaken?' (Cowdale 2003).

The decline of the model is often precipitated by the customer perception that the simulation is no longer providing credible results. It is therefore clearly in the simulation modeller's best interests to have the process and the evidence available to pursued the customer that the model remains 'fit for purpose' (Sadowski 2005).

During the lifetime of the model, there may however be opportunities to collect real world data that can be used in model validation. Within the defence domain these opportunities can sometimes occur during the Operational Test and Evaluation

phase of an acquisition programme or during routine in-service training events. While immensely valuable for the validation of the simulation model, trials or training events of this nature can however present their own challenges and analysis issues.

4 MODEL VALIDATION ISSUES

The United States Department of Defense (DoD) have identified the importance of experimentation to exploit the opportunities that the Information Age concepts and technologies offer. The DoD Code of Best Practice (Alberts and Hayes 2002) describes the logic and anatomy of experiments and provides guidance on how to collect and analyses experimentation data.

A dedicated data collection trial will pose two major challenges for the modeler;

1. how should the trial be designed
2. what data should be collected.

Designing the data collection trial will invariably be an iterative process of using the simulation model to help plan the trial and then utilizing the results to aid the validation of the simulation model (Figure 2).

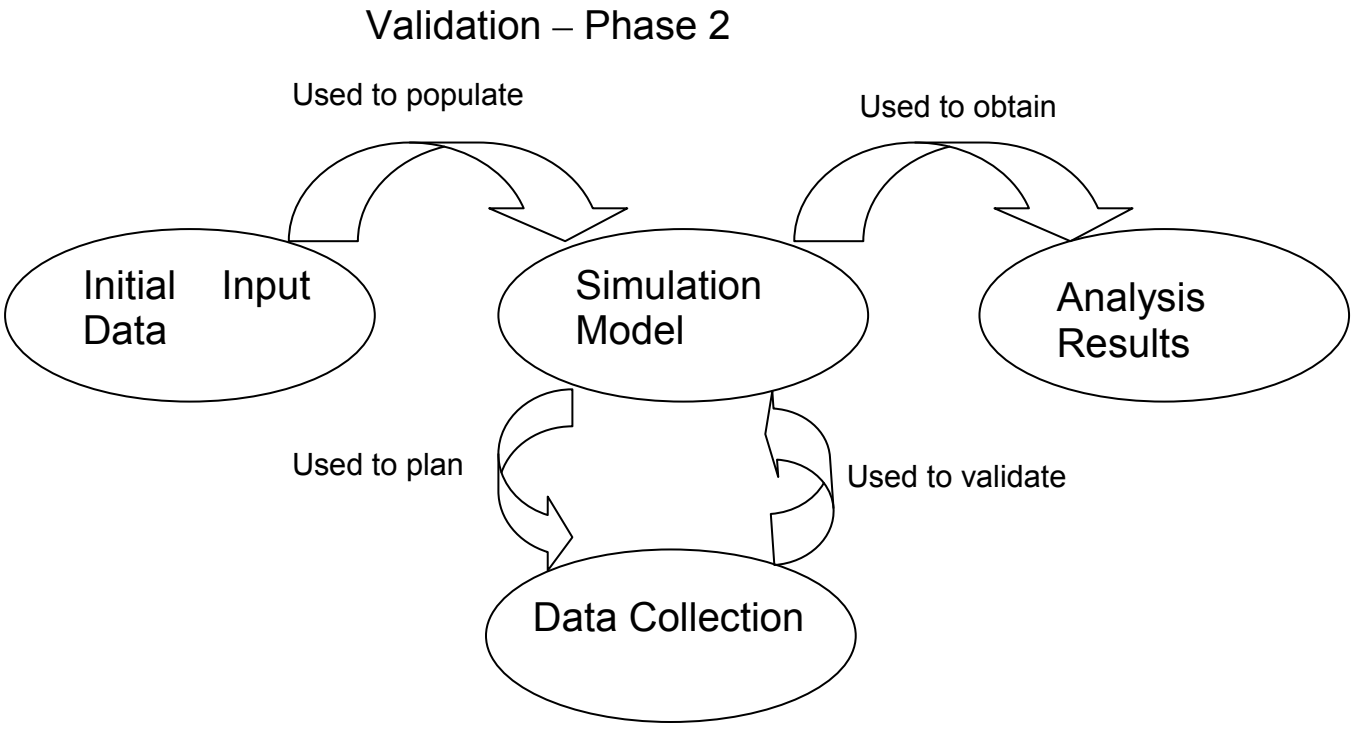

Figure 2: Data Collection For Model Validation

The simulation model can be used to de-risk the trial by predicting likely outcomes and hence identifying data items which need to be collected, and by undertaking sensitivity analysis to ensure the robustness of the proposed trials.

Within the defence environment data collection is often very expensive. Weapon system hardware can be very expensive even before being modified to collect in-flight data. Flight hours are expensive and trials facility time is often difficult to arrange and expensive to facilitate. Consequently the number of data collection opportunities are invariably very limited. The resultant small sample size will provides some particular challenges for the analyst when it comes to establishing the statistical validity of analysis and validation results.

4.1 Trials Design

The data collection trial should be planned using a formal Design of Experiments (DoE) methodology. The complexity of the system under test will however invariably result in a large number of potential factors for both the controllable and the variable input parameters.

While the controllable factors (such as aircraft type, aircraft altitude, speed etc) will present the key DoE decisions (in terms of selecting individual test point parameters), the variable factors may often prove the greater challenges in terms of data collection and risk mitigation.

Variable factors within the physical domain (such as weather, time of day, software version) can generally be measured and recorded, while variable factors in the cognitive domain (such as operator experience, stress) can either be difficult to measure or unethical to test.

Usually the sample size for the trial will be very limited, and hence the number of controllable factors that can be explored will be significantly constrained. The actual experimental design can be based on a traditional factorial design methodology (Fisher 1926) or variations such as those proposed by Plackett-Burman (Plackett 1946), Box-Behnken (Behnken 1960) or more controversial methods such as Robust Design proposed by Taguchi (Taguchi 1987).

4.2 Data Collection

"Collect everything you can" is always a good trials maxim. However, the instrumentation of platforms and systems is often very expensive and physical data collection can be very manpower intensive, thus constraining the practical data collection potential.

For example a weapon effectiveness trial may wish to have specially instrumented targets, environmental sensors, and video cameras to record the events, and may require suitable resources to identify and collect fragments from the weapon and debris from the target post-event. Some of the data collection activities may also be time-critical to ensure that the site or relevant equipments have not become corrupted or contaminated.

Collecting data from the cognitive domain (for example, understanding why participants undertook a particular action) will probably involve either an interview or a questionnaire approach, and may also require suitable resources to pilot the techniques before the trial. Again, consideration needs to be made as to the timing of any interview or questionnaire to ensure accurate and reasoned accounts are obtained.

The data collection plans needs to be integrated with the data analysis plan, to ensure that what is being collected will be of the required depth and fidelity to enable successful analysis to be undertaken.

The training requirement for the trials team should not be underestimated. It is essential that the team collecting the data have a clear understanding of the data requirements, but also have sufficient understanding of the end use of the data to be in a position to identify anomalous events and data items.

4.3 Unexpected Occurrences

While the purpose of the trial may be to provide validation data for the simulation model, the trial itself may present issues that question the overall understanding or credibility of the system under test. For example if a sensor fails to acquire the designated target, or a missile fails to guide correctly, or a weapon fails to detonate.

In these circumstances the analyst will be faced with a number of options;

1. Continue with the trial
2. Terminate the trial
3. Modify the trial plan.

For each of these options the analyst will need to have developed a decision criteria for how to proceed. While a sequential analysis approach (Wald 1946) can aid in the decision as to whether the trial should continue or be terminated, there may be value in continuing the trial, but focusing on collecting data to aid in the understanding of the observed events rather than in the validation of the simulation model.

As a simple example, if 10 weapons with an expected reliability of 90% are allocated to the trial, 3 of the weapons could fail on the trial and the reliability would still be within the 95% statistical confidence limits for a binomial distribution. Clearly if all 10 weapons are successful then more data is collected, but the trials may need to be planned to be adaptable for up to 3 system failures.

5 REAL WORLD EXAMPLES

Data is a series of observations or measurements, which when placed into context becomes Information. The addition of judgement to the interpretation of the Information constitutes Knowledge. Whilst inputs to the simulation model will consist of data, knowledge of the system under test and of the collection process will invariably be required to ensure the data is relevant for the validation activity.

While utilizing an automated method for data collection is obviously advantageous, there will usually be some data elements (particularly in the cognitive domain) that will need to be collected manually. However even if the data collection process is fully automated, there is significant benefit in the analysts observing the data being collection to able to put the key events into context. Understanding the importance of observing how the data is collected is nothing new. On 25th May 1694 Sir Issac Newton in a letter to Nathaniel Hawes wrote; "If instead of sending the observations of able seamen to able mathematicians on land, the land would send able mathematicians to sea, it would signify much more to the improvement of navigation and the safety of men's lives and estates on the element", and these sentiments are equally applicable today.

When the trial is undertaken, there will invariably be events that were not identified in the original data collection plan; for example, the unavailability of a member of the team, unexpected weather or the failure of some pieces of equipment. To mitigate against these occurrences a good data collection plan will have included a detailed risk assessment of potential events that would affect the data collection activity and devised a 'Go/No-Go' checklist of 'essential' elements of the trial. Even with a detailed risk assessment, there is still the possibility that an unforeseen event will occur, in which case the analyst will been to have an in-depth knowledge of both the system under test and the data requirements to maintain the flexibility of actions to deal with any eventuality.

Without a detailed knowledge of the system under test and the data collection system, there may be many pitfalls awaiting the analyst. Some of the most common are:

1. Undertaking the wrong test – either through the system under assessment being too immature for the test or the test not producing data in a useful form for subsequent analysis.
2. Undertaking the test in an unrepresentative environment – either through the use of artificial scenarios or as a direct result of the data collection and instrumentation processes. In the most extreme cases where morale and stress can have a significant impact, the results the differences from military trials and real combat can be an order of magnitude different (Rowland 2006), but even in routine trials, the preparation of the equipments and the training of the participants can significantly affect the outcome.
3. Lack of resources – either through budget constraints or through the lack of time and effort available to collect and process the available data.
4. Not obtaining the right data – either through not recording and collecting the correct data, or collecting the data in an inappropriate format.
5. Biased data – either through the subjective nature of questionnaire style data collection, or as a result of the failure to identify and record anomalous events.
6. Reliance on a single source of data – while the failure of the primary data sources may well negate the test event, by the establishment of a backup data collection process (even with lower data fidelity) it may still be possible to provide useful analysis to support the validation process.
7. Data archiving – either not archiving all of the relevant data for later review and audit, or not providing suitable meta-tagging of the data to provide timely access to the appropriate data items if and when required for follow-on analysis.

All of these types of pitfalls are potentially avoidable. However, in practice they all present real risks to the trial, as the analyst will invariably be constrained in resources and time in the build-up to the trials event and hence will have limited ability to undertake all of the desirable mitigation actions. In these circumstances, it is essential that the analyst utilizes there own experience and those of other practitioners to ensure that previous Lessons Identified can be converted into Lessons Learnt.

6 CONCLUSION

If you are in a position to collection data for model validation;

1. Think very hard about what you want.
2. If in doubt collect it.
3. Make sure you are collecting what you think you are collecting.
4. Ensure you document what you collected and what you didn't.
5. If possible confirm via two sources.
6. Remain flexible.

REFERENCES

Alberts, D., R. Hayes. 2002. Code of Best Practice for Experimentation. CCRP Publication Series.

Balci, O. 1994. Validation, Verification, and testing techniques Throughout the life Cycle of a Simulation Study. In *Proceedings of the 1994 Winter Simulation Conference*, eds. J. D. Tew, S. Manivannan, D. A. Sadowski and A. F. Seila, 215–220. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Box G., D. Behnken. 1960. Some new three level designs for the study of quantitative variables. *Technometrics*. 2: 455-475.

Brown, L. D., T. T. Cai, and A. Dasgupta. 2002. Confidence Intervals For a Binomial Proportion and Asymptotic Expansions. *The Annals of Statistics* 30(1): 160 - 201.

Cowdale, A. C. 2003. December. Simulation modelling in the support of emergency fire-fighting in Norfolk. In *Proceedings of the 2003 Winter Simulation Conference*, eds. S. E. Chick, P. J. Sanchez, D. M. Ferrin and D. J. Morrice, 1707-1710. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Cowdale, A. C. 2006. Lessons Identified from Data Collection for Model Validation. In *Proceedings of the 2006 Winter Simulation Conference*, eds. L. R. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol and R. M. Fujimoto, 1280-1285. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Cowdale, A. C. 2005. Sports Analogy for Modelling of Combat in the Air Domain. In *Proceedings of the 2005 Winter Simulation Conference*, eds. M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, 1123-1128. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Fisher, R. A. 1926. The Arrangement of Field Experiments. *Journal of the Ministry of Agriculture of Great Britain*. 33: 503-513.

Lada, E. K., N. M. Steiger, M. A. Wagner, and J. R. Wilson. 2005. Introduction to modeling and generating probabilistic input processes for simulation. In *Proceedings of the 2005 Winter Simulation Conference*, eds. M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, 41-55. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Law, A. M., and D. W. Kelton. 2000. *Simulation Modeling and analysis*. 3rd Edition. New York: McGraw-Hill.

Ministry of Defence UK. 2008. Defence Test & Evaluation Strategy. Available via < www.mod.uk >.

NATO. 2002. NATO Code of best practice for C2 assessment. CCRP Publication Series.

Plackett, R, L,. and J. P. Burman. 1946. The Design of Optimum Multifactorial Experiments. *Biometrika* 33(4): 305-325.

Rowland, D. 2006. The Stress of Battle. The Stationary Office.

Sargent, R. G. 2003. Verification and validation of simulation models. In *Proceedings of the 2003 Winter Simulation Conference*, eds. S. E. Chick, P. J. Sanchez, D. M. Ferrin and D. J. Morrice, 37-48. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Sadowski, D. A. 2005. Tips for successful practice of simulation. In *Proceedings of the 2005 Winter Simulation Conference*, eds. M. E. Kuhl, N. M. Steiger, F. B. Armstrong and J. A. Joines, 56-61. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Taguchi, G., and S. Konishi. 1987. Orthogonal Arrays and Linear Graphs. Dearborn, MI, ASI press.

Wald, A. 1946. *Sequential Analysis*. J Wiley and Sons Inc.

AUTHOR BIOGRAPHIES

ALAN COWDALE is the senior scientific advisor at the Air Warfare Centre in Lincolnshire, UK. He holds a BSc in Mathematics from the University of Southampton and an MSc in Operational Research from the University of Birmingham. He has undertaken a wide range of simulation projects for both the Royal Navy and Royal Air Force in connection with assessing weapon system performance and effectiveness. His email is <accowdale@dstl.gov.uk>.